

Return of fear after retrospective inferences about the absence of an unconditioned stimulus
during extinction.

An K. Raes

Jan De Houwer

Bruno Verschuere

Rudi De Raedt

Ghent University, Belgium

Corresponding author:

An Raes

Ghent University, Faculty of Psychology

Department of Experimental-Clinical and Health Psychology

Henri Dunantlaan 2

B-9000 Ghent

Belgium

Tel: +32 (0)9 264 94 42

Fax: +32 (0)9 264 64 89

Email: An.Raes@UGent.be

Abstract

We examined whether the effect of an extinction phase can be influenced retrospectively by information about the cause of the absence of the unconditioned stimulus (US) during that phase. Participants were subjected to a differential fear conditioning procedure, followed by an extinction procedure. Afterwards, half of the participants were presented with information about a technical failure, which explained why the US had been absent during the extinction phase. The other participants received information that was unrelated to the US. During a subsequent presentation of the target conditioned stimulus (CS), only the former group of participants showed renewed anticipatory skin conductance responding and a return of US expectancy. The results are in accordance with a propositional account of associative learning and highlight the importance of retrospective reasoning as a cause of relapse after exposure therapy.

Keywords: human fear conditioning; extinction; propositions; retrospective inferences; return of fear; electrodermal responding

Return of fear after retrospective inferences about the absence of an unconditioned stimulus during extinction.

The decrease in conditioned responding following unpaired presentations (CS only) of a previously reinforced stimulus (CS-US) is referred to as extinction. The fact that extinction has been studied extensively can be explained by its relevance to clinical practice (Hermans, Craske, Mineka, & Lovibond, 2006). Clinical procedures such as exposure treatment can be seen as analogues of extinction procedures. Therefore, broadening our understanding of extinction can help us to further refine exposure.

Several authors have focused attention on the cognitive processes that underlie extinction effects (e.g., Bouton, 1993, 2002; Hermans, et al., 2006; Lovibond, 2004). Most theories postulate the existence of associations in memory and explain extinction in terms of a change in the strength of the association between the CS and the US (Rescorla & Wagner, 1972) or through the emergence of a new inhibitory association (Bouton, 1993, 2002). Another class of theories describes the extinction process in terms of the formation and evaluation of propositions (De Houwer, 2009; Lovibond, 2003; Mitchell, De Houwer, & Lovibond, 2009). According to the propositional account, the extinction procedure allows participants to verify that the CS is no longer followed by the US. In most cases, this observation will lead to the proposition that the CS is no longer a good predictor of the US and that future presentations of the CS will not be followed by the US. The latter proposition is thought to be responsible for the behavioural extinction effect.

Both associative and propositional models can explain that extinction effects tend to be context specific, in the sense that a context change after extinction readily leads to renewal. When the extinction trials take place in another room than that in which the original CS-US pairings occurred, they have little effect on the CR that the CS evokes in the original context (Vansteenwegen, et al., 2005). Association formation models account for this by assuming

that the associations that are formed during the extinction procedure and that produce the extinction effect are active only in the context in which they are formed (Bouton, 1993, 2002). On the other hand, propositional models account for the context specificity of extinction by postulating that people might not use the extinction trials to infer that there has been a *general* change in the CS-US relation (Lovibond, 2004). That is, if there are reasons to believe that the absence of the US during the extinction phase was due to a third factor that was present only during the extinction trials, they might doubt the validity of a general inference, and allocate the change in the CS-US relation to that third factor. A change in context might signal the presence of such a third factor and thus raise doubts about whether the CS-US relation will be different when the extinction context is removed.

Interestingly, according to propositional models, the truth of propositions can also be revised retrospectively, that is, on the basis of new information that is given after the initial proposition has been formed. In human causal learning literature, there are myriad demonstrations of the effect of retrospective revaluation (Dickinson & Burke, 1996; Mitchell, Killedar, & Lovibond, 2005). In the context of fear conditioning, Davey's research on US revaluation supports the notion of retrospective revaluation (see Davey, 1992; Davey, 1997, for overviews). Davey and colleagues demonstrated that a change in the evaluation of the US can lead to a substantial change in the CR evoked by the CS. Importantly, the revaluation of the US can occur completely independent of CS-US occurrences and can be installed through direct experience (White & Davey, 1989) or instructions (Davey & McKenna, 1983). These findings demonstrate that a retrospective change in the *content* of the US representation can affect subsequent responding to the CS. However, until now, it has not been examined whether retrospective inferences on the absence of the US might lead to a return of CR or US expectancy after extinction

The aim of the present study is to examine the effect of retrospective inferences on US absence after extinction. Similar to the work of Davey, this type of retrospective revaluation pertains only to the US and not to the CS. However, in contrast with Davey, the retrospective revaluation does not involve the content of the US (e.g., the degree to which it is aversive or threatening) but refers to the absence of the US during a past extinction phase. Based on the propositional approach, we would expect that also in this case retrospective inferences will lead to a change in the evaluation of the CS-US relationship. More specifically, we expect that an explanation for the absence of the US in a past extinction phase will reinstall US expectancy and conditioned responding on subsequent CS presentations.

It is important to examine the effect of retrospective inferences on the absence of the US after extinction because similar processes could operate in a clinical setting. For example, people could infer after exposure therapy that the phobic object was not followed by an aversive event during the therapy (e.g., they were not bitten by the dog) because therapy had taken place under specific circumstances (e.g., during summer; people might think, for example, that dogs are relaxed and generally non-aggressive during that time of year). Following the propositional account, this assumption alone could lead to return of fear when those specific circumstances are not longer believed to be present (e.g., it is no longer summer).

Our research question is also theoretically important. Whereas propositional models predict that retrospective inferences can modulate extinction, it is difficult to see how such effects can be explained by associative models. According to these models, learning (i.e., changes in associative strength) can occur only if the CS is physically present (e.g., Rescorla & Wagner, 1972) or when the CS representation is activated by the physical presence of another stimulus that previously co-occurred with the CS (e.g., Dickinson & Burke, 1996). Hence, verbal information about a cause that never physically co-occurred with the CS or US

should not influence the strength of the CS-US association and should thus not influence the extent to which the CS evokes a CR.

Because of practical and ethical concerns, we examined this hypothesis in a laboratory setting. Participants were subjected to a differential conditioning procedure, in which one of two cues (CS+) was paired with the US during acquisition, whereas another cue (CS-) was never paired with the US. The effect of conditioning was measured through indexing both anticipatory skin conductance responses (SCRs) and US expectancy for both CSs. Both groups also experienced an extinction phase during which both CSs were presented without the US. After the extinction phase, one group of participants was told that there had been a technical problem because of which the US had not been presented for some time, but that the problem had now been solved by the experimenter. This information allows participants to infer that the absence of the US during extinction was due to that technical failure rather than to a change in the relation between the CS and US. Because the technical problem had been solved, they should expect the US after the next presentation of the CS+ and thus show a conditioned response to the CS+. We refer to this group of participants as the retrospective inference group (RIG).

To exclude that changes in extinction effects were due to the mere fact of giving an instruction after the extinction trials, another group of participants (control group; CG) was provided with control information that did not refer to the US. These participants were told that there had been a technical problem (no reference to US) which had now been solved by the experimenter. After the presentation of this information, both the RIG and the CG were presented with a CS- presentation, followed by a non-reinforced CS+ presentation (test phase). The fixed order served to prevent order-effects, with the CS- eliciting a stronger response when preceded by an unexpectedly non-reinforced CS+ (Lovibond, 2003; as in Vervliet, Vansteenwegen, Baeyens, Hermans, & Eelen, 2005).

Based on the propositional account, we expected that the RIG, but not the CG, would show a return of CR and US expectancy for the CS+ in the test phase.

Method

Participants

Thirty-eight undergraduate students (seven men) at Ghent University participated in this experiment in exchange for course credits or eight Euros. Mean age was 18.84 ($SD = .97$). Four participants were excluded from further data-analysis because their ratings on the trait version of the Spielberger State-Trait Anxiety Inventory (STAI-T) (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983) exceeded the sample mean by more than 2.5 SD. One participant was excluded because his score on the Beck Depression Inventory (BDI; Beck, Steer, & Brown, 1996) exceeded the sample mean by more than 2.5 SD. Allocation to the RIG and CG was randomised. Eighteen participants were part of the RIG group, 15 participants were part of the CG. Mean age was similar in the RIG ($M = 18.72$, $SD = 0.89$) and CG ($M = 18.67$, $SD = 0.90$), $t < 1$. Also, there was a similar gender distribution in both groups (three and four men respectively in the RIG and CG groups), $\chi^2 = 0.49$, ns . BDI (RIG: $M = 5.28$, $SD = 3.03$; CG: $M = 4.07$, $SD = 2.60$) and STAI-T (RIG: $M = 35.61$, $SD = 4.55$; CG: $M = 33.13$, $SD = 8.00$) scores did not differ significantly between groups, $t(31) = 1.22$, ns (BDI) and $t(31) = 1.12$, ns (STAI-T).

Material

Experimental stimuli. All stimuli were presented centrally on a computer monitor with a resolution of 1024 x 768 pixels. The background colour was black. Two coloured slides (335 x 312 bitmap files) served as conditioned stimuli. One slide was purple, the other one was green. One of these slides was sometimes (i.e., during acquisition) followed by the US (CS+) while the other one (CS-) was never followed by the US. The allocation of the

slides to the function of CS+ and CS- was counterbalanced. A 170 ms 95 dB(A) white noise served as US. The noise was presented binaurally with Philips headphones. The volume of the white noise was checked by technical staff before the start of the study. Using a sound level meter (Brüel and Kjær's Type 2250; Nærum, Denmark) the dB(A) level of the white noise was measured in the ear pads of the headphones.

US expectancy ratings. US expectancies were assessed retrospectively for the CS+ and the CS- separately (as in Vansteenwegen et al., 2005; Vervliet et al., 2005). Participants were asked to indicate the evolution of their US expectancies for the CS+ and the CS- on separate graphs. The X-axis of this graph represented the time course of the experiment, with nine separate rating moments (first and second baseline trial; beginning, middle, and ending of acquisition; beginning, middle, and ending of extinction; test trial). The Y-axis represented the strength of US expectancy on a scale ranging from 0 (I do not expect the US at all) to 10 (I expect very strongly that the US will be presented).

Psychophysiology. Skin conductance responses were recorded with a Coulbourn Lablinc V, which was gated to a PC through a Scientific Solutions DMA card. Signals were digitised through customised software (Psychophysiological Recording; PSPHR). Skin conductance was recorded with standard Ag/AgCl electrodes (0.8 cm diameter) filled with KY-jelly. The electrodes were attached on the thenar and hypothenar eminences of the non-preferred hand, which was first cleaned with tap water. The signal was measured using a constant voltage (0.5 V) coupler, and digitised at 10 Hz.

The recorded data were analysed off-line with Psychophysiological Analysis (PSPHA) (de Clercq, Verschuere, de Vlieger, & Crombez, 2006). For each trial, SCR (in μ S) was calculated by subtracting the mean value of a baseline period (2 s before CS onset) from the highest amplitude in a 1-8 s time window after CS onset (Pineles, Orr, & Orr, 2009). To account for individual differences in responsivity (Lykken & Venables, 1971), amplitudes

were range corrected using the largest measured response for that participant during the entire experiment (including two US trials, one at the start of the experiment and one at the end of the experiment). Finally, in order to normalise the data (Dawson, Schell, & Fillion, 2000), range corrected amplitudes were square root transformed prior to further analysis.

Subjective ratings. US valence and US pain were assessed by visual analogue scales (VASs) ranging from 0 (not at all unpleasant/painful) to 10 (very unpleasant/ painful). We included a post-experimental check of the believability of the experiment including two questions. First, participants were asked to what extent they believed the instructions in general by indicating a number ranging from 0 to 10 (with 0: the instructions were not at all believable; 10: the instructions were very believable). Second, participants were asked to indicate which instructions, if any, they found less believable than the other instructions.

Procedure

Upon entrance in the experiment room, participants were asked to complete the STAI-T and the BDI. Subsequently, the experimenter showed the participant the apparatus that was used to record skin conductance responses. All participants were given the following instructions:

“During this experiment, your skin conductance responses will be measured. You will be asked to sit in a separate room that is adjacent to this room and electrodes will be attached to your non-dominant hand. The apparatus here will record your skin conductance responses and these will be visible online to the experimenter. This kind of measurement is very delicate. Therefore, the experimenter will repeatedly check the apparatus during the experiment. At these moments, a message will appear on your computer screen. You will be asked to pause for some moments, until the experimenter confirms that you can continue the experiment.”

These instructions allowed us to prepare participants for repeated technical check-ups which, in turn, increased the truth-value of the crucial information that was given at the end of extinction. All participants were presented with three technical check-up messages. Two messages were presented in the middle of the acquisition and extinction phases, respectively. These moments were chosen to prevent that the presentation of the messages would co-occur with a change in contingencies. The third message, which included the crucial experimental manipulation, was provided at the end of the extinction phase.

After the initial instructions, participants were taken to the adjacent room. The experimenter asked the participants to clean their hands with tap water and to take place in front of the computer screen. Then, the experimenter attached the electrodes. Based on ethical considerations, participants were confronted with the US before the start of the conditioning task. They could resign from participation at any time without consequences.

Before the start of the baseline phase, participants were instructed that they would be repeatedly presented with two coloured slides, one of which, during later stages of the experiment, would be paired with the white noise (US). Participants were asked to attend to the contingencies between the coloured slides and the noise and were informed that they would be asked to report US expectancies for both slides after the task.

Baseline phase. Two CS+ and two CS- pictures were presented in semi-randomised order in that not more than two consecutive trials of CS+ or CS- were presented. Each trial started with a 3 s presentation of a fixation cross which announced the CS. The CS was presented for 8 s. The inter-trial interval varied between 10 s and 18 s (average of 14s).

Acquisition phase. An instruction on the screen announced that the loud white noise (US) could be presented at the offset of one of the two CS pictures. Participants were asked to attend to the contingencies between the pictures and the noise for the rest of the experiment. Sixteen acquisition trials were included. Both the CS+ and the CS- were presented eight times

in semi-randomised order (two blocks of eight randomised trials, each consisting of four CS+ and four CS- trials). The CS+ was consistently followed by the US during this phase. The US was presented at CS offset. Apart from this, trial course was identical to baseline.

In the middle of acquisition (after four CS+ and four CS- trials), the acquisition phase was shortly interrupted by a first technical check-up message on the computer screen.

Participants were asked to pause for some moments, as the experimenter would check the apparatus. After half a minute approximately, the experimenter entered the room and informed the participant that the measurement was going well. Participants were instructed to continue the experiment after the experimenter had closed the door of the room.

Extinction phase. Participants received no extra instructions before the start of this phase, nor were they alerted in any sense that a new experiment phase had started. All participants were presented with 10 CS+ and 10 CS- trials in semi-randomised order (two blocks of 10 randomised trials, each consisting of five CS+ and five CS- trials). No USs were presented anymore. Except for the absence of the US, trial course was identical to acquisition. Similar to acquisition, a technical check-up message appeared on the computer screen at the middle of this phase (after five CS+ and five CS- trials). The interruption by the experimenter was identical to the interruption that had taken place during acquisition.

At the end of extinction, a third technical check-up message was presented. In both the control group (CG) and the retrospective inference group (RIG), the experimenter announced via the intercom that there was a technical failure and participants were asked to refrain from continuing the experiment. After one minute approximately, the experimenter entered the participants' room. Participants in the CG were told that there had been a technical disturbance which was now solved. Participants were asked to continue the experiment. The RIG was told that there had been a technical disturbance. Through this disturbance, no noise

had been presented for some time. However, the disturbance was now solved and the experiment could continue.

If a participant asked whether it made sense to continue the experiment after a technical failure, the experimenter replied that this was one of the first times that the experiment was run. Therefore, completion of the experiment was useful to ensure that there were no further problems with the script or with the technical equipment and to check whether the data were recorded appropriately. This was relevant for one participant (in the RIG) only.

Test phase. Participants were presented with one CS- trial, followed by one non-reinforced CS+ trial.¹ At the end of this phase, participants were asked to rate the white noise (US) on valence and painfulness using paper and pencil.

US-expectancy ratings. After the conditioning task, the experimenter entered the participants' room to remove the electrodes. Participants were taken to the adjacent room and asked to complete US expectancy ratings for the CS+ and the CS- (separately) using a graph with nine fixed time points (see above) (cf. Vansteenwegen, et al., 2005; Vansteenwegen, et al., 2006; Vervliet, et al., 2005). The experimenter explained in detail what each time point was referring to. With regard to the last time point, each participant was reminded specifically that this trial had taken place after the last technical check-up.

When the US expectancy ratings were completed, all participants were given a questionnaire to assess the extent to which they had believed (a) the instructions in general and; (b) the specific instruction that had been given after the extinction phase. Participants were told that it was important for the experimenter to know what participants had been thinking during the experiment, because this could have influenced their data pattern. All participants were debriefed at the end of the experiment.

Results

US expectancies

Not all trials were analysed. We chose to analyse only four crucial moments: the first baseline trial (ba1), the last acquisition trial (acq3), the last extinction trial (ext3), and the test trial. US expectancy ratings were subjected to a 4 (Phase: ba1, acq3, ext3, test) x 2 (CS: CS+, CS-) x 2 (Group: CG, RIG) analysis of variance (ANOVA) with phase and CS as within-subject factors and group as a between-subjects factor. This analysis resulted in significant main effects of phase, $F(3,29) = 13.24, p < .001$, partial $\eta^2 = .58$, and CS, $F(1,31) = 174.41, p < .001$, partial $\eta^2 = .85$, and yielded a significant Phase x CS interaction, $F(3,29) = 65.91, p < .001$, partial $\eta^2 = .87$. All foregoing effects subsumed under a significant three-way interaction, $F(3,29) = 3.37, p = .03$, partial $\eta^2 = .26$. We followed up on this interaction by performing 2 (CS) x 2 (Group) ANOVA's for each experiment phase. During baseline, there were no significant main or interaction effects concerning CS or group, F 's < 1.55 . At the end of acquisition, US expectancies were significantly higher for the CS+ ($M = 8.76, SD = 2.09$) than for the CS- ($M = 1.27, SD = 1.84$), as illustrated by a main effect of CS, $F(1,31) = 265.20, p < .001$, partial $\eta^2 = .90$, but there were no main or interaction effects involving group, F 's < 1 . At the end of extinction, US expectancies for the CS+ ($M = 5.45, SD = 2.10$) were still significantly higher than those for the CS- ($M = 1.76, SD = 2.24$), $F(1, 31) = 70.87, p < .001$, partial $\eta^2 = .70$. Importantly, however, CS+/CS- differentiation was significantly smaller than during acquisition, $F(1, 31) = 76.07, p < .001$, partial $\eta^2 = .71$. Analysis of the extinction data did not yield significant effects involving group, F 's < 1 .

In the test phase, the significant main effect of CS, $F(1, 31) = 72.87, p < .001$, partial $\eta^2 = .70$, was qualified by a significant CS x Group interaction, $F(1, 31) = 9.00, p = .005$, partial $\eta^2 = .26$. Although both groups displayed higher US expectancies for the CS+ than for

the CS-, discrimination was larger in the RIG (CS+: $M = 7.28$, $SD = 3.10$; CS-: $M = 1.17$, $SD = 1.47$), $t(17) = 7.80$, $p < .001$, $d = 1.97$, than in the CG (CS+: $M = 4.93$, $SD = 2.40$; CS-: $M = 2.00$, $SD = 2.42$), $t(14) = 4.32$, $p < .001$, $d = 1.12$ (see Figure 1).

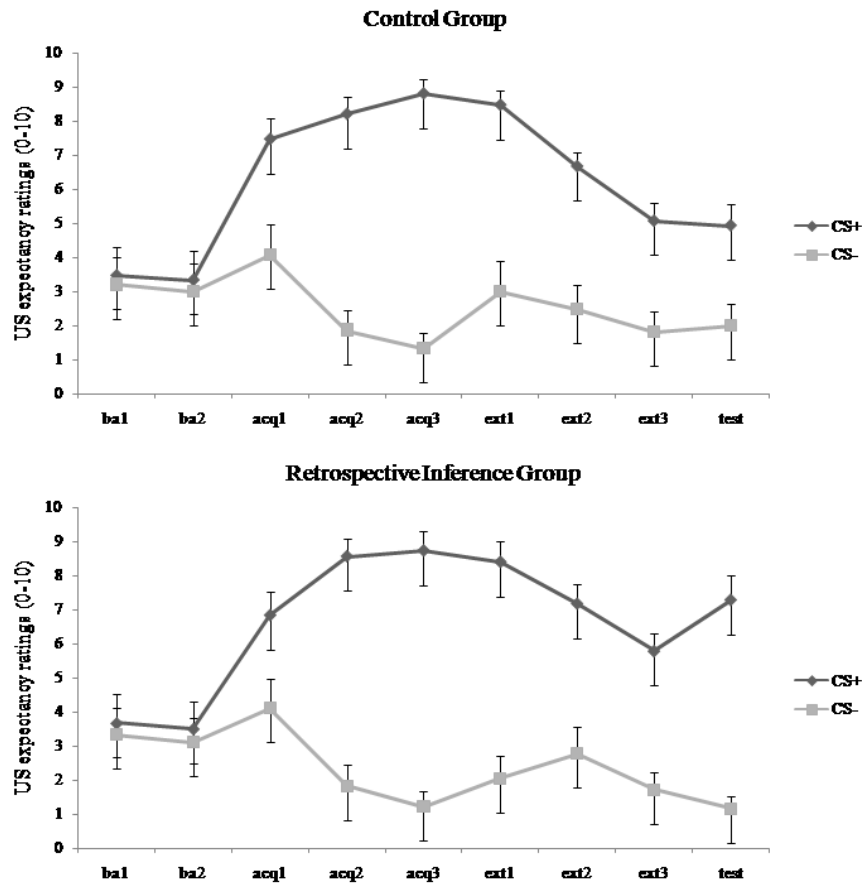


Figure 1. US expectancies for the CS+ and the CS- during baseline, acquisition, extinction, and test, represented separately for the control group (CG) and the retrospective inference group (RIG).

To ensure that there was a renewal effect in the RIG but not in the CG, a between-phase (extinction versus test) comparison was performed for the CS+ and the CS- separately in both groups. In the RIG, there was a significant increase in US expectancy from the last extinction trial to the test trial for the CS+, $t(17) = 2.39$, $p = .03$, $d = 0.60$. For the CS-, there was a marginally significant decrease, $t(17) = 1.97$, $p = .07$, $d = 0.54$. In the CG, US expectancies for both the CS+ and the CS- did not differ significantly between the last extinction trial and the test trial, t 's < 1 .

Skin Conductance Responding

Three participants were excluded from this analysis because of zero responses on more than 80 % of the trials. Analyses of SCRs were performed on thirty subjects, of whom 16 were part of the RIG and 14 were part of the CG.

A 4 (Phase: baseline, acquisition, extinction, test) x 2 (CS: CS+, CS-) x 2 (Group: CG, RIG) ANOVA on SCRs with phase and CS as within-subject factors and group as a between-subjects factor was conducted. Because of violations of sphericity, Greenhouse-Geisser statistics are reported for this analysis. The main effects of phase, $F(3,26) = 6.56, p = .005$, partial $\eta^2 = .19$, and CS, $F(1,28) = 30.98, p < .001$, partial $\eta^2 = .53$, reached significance. There were also significant interactions of phase and CS, $F(1,28) = 7.08, p = .003$, partial $\eta^2 = .20$, and of CS and group, $F(3,26) = 6.84, p = .01$, partial $\eta^2 = .20$. The three-way interaction effect almost reached significance, $F(3,26) = 3.08, p = .06$, partial $\eta^2 = .10$. Because of the relevance to our strong a priori hypotheses on differences between the RIG and the CG, we followed up on this interaction by performing CS x Group ANOVA's for each experiment phase separately.

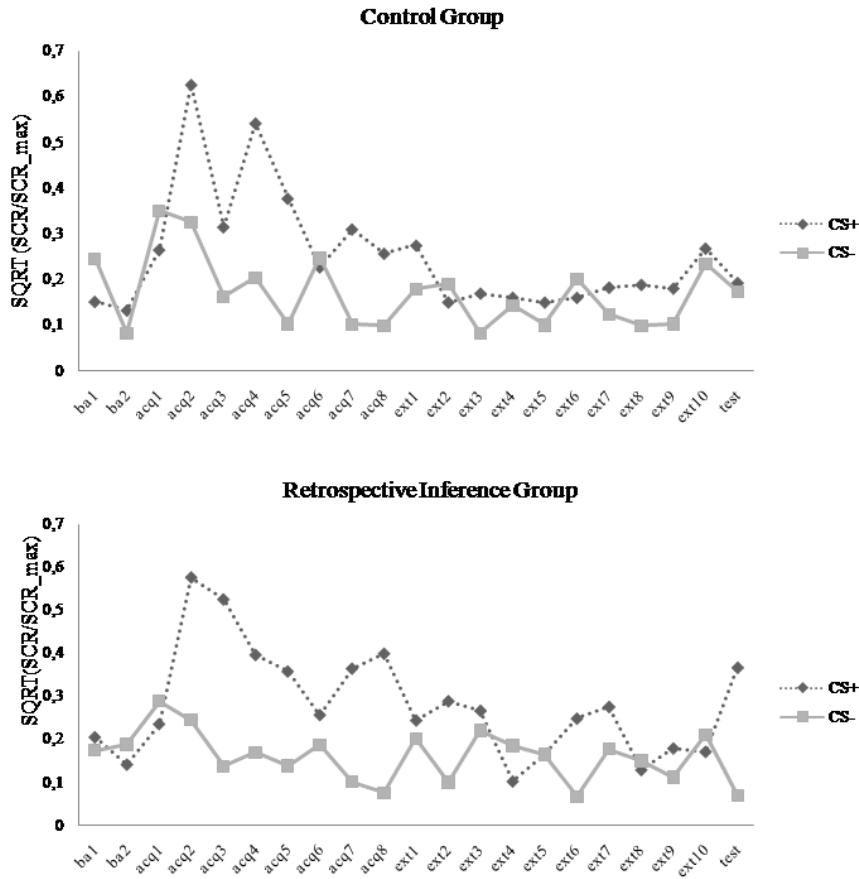


Figure 2. Skin conductance responses for the CS+ and the CS- during baseline, acquisition, extinction, and test, represented separately for the for the control group (CG) and the retrospective inference group (RIG).

During baseline, no significant main or interaction effects were detected, F 's < 1 . During acquisition, a significant main effect of CS, $F(1,28) = 64.25$, $p < .001$, partial $\eta^2 = .70$, indicated that the CS+ elicited significantly larger SCRs ($M = 0.38$, $SD = 0.18$) than the CS- ($M = 0.18$, $SD = 0.12$). There were no significant effects involving group during this phase, F 's < 1.36 . During extinction, CS+/CS- discrimination disappeared (CS+: $M = 0.20$, $SD = 0.16$; CS-: $M = 0.15$, $SD = 0.11$), with the main effect of CS being no longer significant, $F(1,28) = 2.50$, ns (see Figure 2). The CS x Group ANOVA for the test phase resulted in a significant main effect of CS, $F(1,28) = 7.43$, $p = .01$, partial $\eta^2 = .21$, which subsumed under a significant interaction of CS and group, $F(1,28) = 5.83$, $p = .02$, partial $\eta^2 = .17$. As shown

in Figure 2, there was no CS+/CS- discrimination in the CG (CS+: $M = .19$, $SD = .26$; CS-: $M = .17$, $SD = .29$), $t < 1$, whereas the CS+ elicited larger SCRs ($M = .37$, $SD = .32$) than the CS- ($M = .07$, $SD = .15$) in the RIG, $t(15) = 3.83$, $p = .002$, $d = 1.09$.

Between-phase (extinction versus test) comparisons confirmed that there was a renewal effect in the RIG, with a marginally significant increase in SCRs for the CS+, $t(15) = 2.04$, $p = .06$, $d = 0.59$, and a marginally significant decrease in SCRs for the CS-, $t(15) = 1.88$, $p = .08$, $d = 0.49$ (see Figure 2). In the CG, SCRs for both the CS+ and the CS- did not differ significantly between extinction and test, $t's < 1$ (see Figure 2).

Subjective Ratings

Thirty-three participants were included in these analyses. Participants rated the US noise as unpleasant ($M = 78.79$, $SD = 26.16$) and moderately painful ($M = 40.91$, $SD = 22.69$). There were no between-group differences, $t's < 0.39$.

Participants assessed the instructions that were given during the experiment as very believable ($M = 8.48$, $SD = 1.89$). These ratings did not differ between the CG and the RIG, $t(31) = 1.37$, ns . One person from the CG (6.7%) and two persons from the RIG (11.1%) reported that they found the final instruction (about the equipment failure during the extinction phase) less believable than the other instructions. There was no significant group difference on this outcome either, Fisher's Exact test, $p = .52$.

Discussion

According to propositional models of associative learning (De Houwer, 2009; De Houwer, Beckers, & Vandenorpe, 2005; Mitchell, et al., 2009), retrospective inferences that give rise to an alternative interpretation of the observed contingencies should lead to a change in conditioned responding. Within the context of extinction, a return of the CR after extinction is expected when people retrospectively attribute the absence of the US during extinction to a third factor that was present only during the extinction phase. In the current experiment, a

technical failure (through which no USs could be presented anymore) served as the third factor. Half of the participants (RIG) were informed about this third factor after extinction and were told that this factor had been removed. As predicted by the propositional account, these participants showed return of fear on a subsequent CS+ presentation. In the other participants (CG), who had received control information that did not refer to the US, the extinction effect remained.

These effects were most pronounced on the US expectancy ratings. A clear increase in US expectancies could be observed from extinction to test in the RIG, but not in the CG. The SCRs parallel these findings, although the effects on this measure were less pronounced. In this context, it should be noted that the power of our study might have been limited due to the small sample size. Nonetheless, although some of our findings were marginally significant, the crucial results were accompanied by moderate to large effect sizes, thus suggesting that our findings are reliable.

The present results are difficult to reconcile with an associative account of fear conditioning (e.g., Rescorla & Wagner, 1972). Even if we assume that verbal information can lead to changes in underlying associations, it is unclear how a verbal message that mentions only the US would lead to changes in CS-US association on a subsequent CS presentation. The contextual associative account of Bouton (1993, 2002) might, however, be able to account for the present findings. For instance, when participants in the RIG were told that a technical error had interfered with the presentation of the US, they might have retrospectively recoded the extinction context. This would then result in an ABA situation where A refers to the context in which there was no technical error (as present during acquisition and test) and B refers to the context in which there was a technical error (as present during extinction). Participants in the CG did not have a reason to recode contexts. Hence, one could say that they were exposed to an AAA situation. Because extinction has more impact on test

performance when it is presented in the same context as acquisition (AAA) then when it is presented in a different context (ABA), conditioned responding during test would be stronger in the RIG than in the CG, as was observed. Although post-hoc explanations in terms of the mental retrospective recoding of contexts are feasible, it is important to realise that they still attribute an important role to propositional processes in extinction. Whereas these alternative explanations maintain the assumption that the knowledge about the CS-US relation is encoded in an associative rather than propositional manner, they do imply that propositional beliefs about the reasons for the absence of the US can (retrospectively) influence the encoding of the context. As such, even a post-hoc associative account of our results needs to acknowledge the impact that retrospective propositional reasoning can have on extinction.

The results of this study indicate that retrospective inferences concerning the absence of the US during extinction may give rise to a return of fear. In a clinical context, this would imply that although cognitions can serve as an important therapeutic tool, they are also powerful in reinstating fear, even after therapy has ended. Therefore, clinicians should deal with these possible retrospective inferences to optimise treatment outcome. One possibility is to anticipate on retrospective inferences during therapy or to provide follow-up sessions after exposure treatment during which these inferences can be addressed. Of course, these clinical implications are very preliminary, as the present study does not provide any direct measurement of fear. Future research should further investigate the role of retrospective inferences in a clinical context.

Several aspects of the present data need further consideration. A first point of discussion concerns the small extinction effects for the US expectancy ratings. One explanation for this might relate to the retrospective nature of US expectancy assessment. As pointed out by Collins and Shanks (2002), retrospective judgements are mostly based on an integrative strategy. As such, participants might have judged US expectancies based on an

integration of the knowledge they obtained during the whole experiment (Collins & Shanks, 2002). In relation to this, it should be noted that measuring US expectancies were questioned after participants had gone through a subsequent phase (test phase). The information that participants were presented with during this phase might have caused them to update their US expectancy judgement for the previous phase.

In spite of this, the extinction phase did have a substantial impact on CS+/CS- discrimination. Although participants still reported higher US expectancy for the CS+ than for the CS- at the end of extinction, this effect was considerably reduced in comparison with the end of acquisition. In addition, reduced extinction effects should have occurred in both groups and can therefore not account for the between-group differences in the test phase.

Another aspect of the present data that warrants further discussion relates to the methodology of the test phase. Participants were presented with a fixed order of presentation (i.e., a CS- trial followed by a non-reinforced CS+ trial) to prevent enhanced responding to the CS- (Lovibond, 2003; such as Vervliet et al., 2005). One could interpret this as installing a bias towards finding a return of the CR. On the other hand, the fixed order was used in both the RIG and the CG and therefore cannot explain the main results of this experiment.

Finally, at the start of the experiment, participants were given explicit instructions about the CS-US pairings that they would see. Although they were not told which stimuli would go together, we did inform them of the fact that one CS would sometimes be followed by a US whereas another CS would not. The reason for this was to ensure that participants would consider one of the CSs (CS-) as a safety signal throughout the whole experiment, in spite of changes in contingencies (e.g., during extinction) or in instructions. However, it would be useful to examine whether our results generalise to a situation in which these explicit instructions are not provided.

In summary, the present results show that extinction effects can be disrupted by retrospective inferences with regard to the CS-US relationship. After extinction, participants in the RIG showed a return of conditioned SCRs and an increase in US expectancy for the CS+. These findings are in accord with a propositional account of learning, which states that associative learning is based on the formation and evaluation of propositions (De Houwer et al., 2005; Mitchell et al., 2009). However, the present findings can also be explained by a contextual associative approach (Bouton, 1993, 2002), but only on the condition that the possibility for propositional reasoning is allowed within this framework. Importantly, these results may help explaining why many patients show a return of fear after successful exposure. Based on this study, we would suggest that therapists pay close attention to inferences that patients make with regard to the absence of the US during exposure, not only during, but also after therapy.

Acknowledgments

An Raes is a Research Assistant of the Research Foundation-Flanders (FWO), Belgium.

Bruno Verschuere is a postdoctoral researcher of Research Foundation-Flanders (FWO), Belgium.

Jan De Houwer is supported by Grants BOF/GOA2006/001 and BOF09/01M00209 of Ghent University

References

- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory manual* (2nd ed.). San Antonio, TX: Psychological Corporation.
- Bouton, M. E. (1993). Context, Time, and Memory Retrieval in the Interference Paradigms of Pavlovian Learning. *Psychological bulletin*, *114*, 80-99.
- Bouton, M. E. (2002). Context, ambiguity, and unlearning: sources of relapse after behavioral extinction. *Biological psychiatry*, *52*, 976-986.
- Collins, D. J., & Shanks, D. R. (2002). Momentary and integrative response strategies in causal judgment. *Memory & Cognition*, *30*, 1138-1147.
- Davey, G. (1997). *Phobias : a handbook of theory, research and treatment*. Chichester: Wiley.
- Davey, G. C., & McKenna, I. (1983). The effects of postconditioning revaluation of CS and UCS following Pavlovian second-order electrodermal conditioning in humans. *The Quarterly journal of experimental psychology. B, Comparative and physiological psychology*, *35 (Pt 2)*, 125-133.
- Davey, G. C. L. (1992). Classical-Conditioning and the Acquisition of Human Fears and Phobias - a Review and Synthesis of the Literature. *Advances in Behaviour Research and Therapy*, *14*, 29-66.
- Dawson, M. E., Schell, A. M., & Fillion, D. L. (2000). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary & G. G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 200-224). Cambridge: Cambridge University Press.
- de Clercq, A., Verschuere, B., de Vlieger, P., & Crombez, G. (2006). Psychophysiological analysis (PSPHA): a modular script-based program for analyzing psychophysiological data. *Behavior research methods*, *38*, 504-510.
- De Houwer, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning & behavior : a Psychonomic Society publication*, *37*, 1-20.
- De Houwer, J., Beckers, T., & Vandorpe, S. N. (2005). Evidence for the role of higher order reasoning processes in cue competition and other learning phenomena. *Learning & Behavior*, *33*, 239-249.
- Dickinson, A., & Burke, J. (1996). Within-compound associations mediate the retrospective revaluation of causality judgements. *Quarterly Journal of Experimental Psychology Section B-Comparative and Physiological Psychology*, *49*, 60-80.
- Hermans, D., Craske, M. G., Mineka, S., & Lovibond, P. F. (2006). Extinction in human fear conditioning. *Biological psychiatry*, *60*, 361-368.
- Lovibond, P. F. (2003). Causal beliefs and conditioned responses: retrospective revaluation induced by experience and by instruction. *Journal of experimental psychology. Learning, memory, and cognition*, *29*, 97-106.
- Lovibond, P. F. (2004). Cognitive processes in extinction. *Learning & Memory*, *11*, 495-500.
- Lykken, D. T., & Venables, P. H. (1971). Direct measurement of skin conductance: a proposal for standardization. *Psychophysiology*, *8*, 656-672.
- Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *The Behavioral and brain sciences*, *32*, 183-198; discussion 198-246.
- Mitchell, C. J., Killedar, A., & Lovibond, P. F. (2005). Inference-based retrospective revaluation in human causal judgments requires knowledge of within-compound relationships. *Journal of experimental psychology. Animal behavior processes*, *31*, 418-424.

- Pineles, S. L., Orr, M. R., & Orr, S. P. (2009). An alternative scoring method for skin conductance responding in a differential fear conditioning paradigm with a long-duration conditioned stimulus. *Psychophysiology*, *46*, 984-995.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning. Variations in the effectiveness of reinforcement and non-reinforcement. . In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory*. (pp. 64-99). New York: Appleton.
- Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Vansteenwegen, D., Hermans, D., Vervliet, B., Francken, G., Beckers, T., Baeyens, F., & Eelen, P. (2005). Return of fear in a human differential conditioning paradigm caused by a return to the original acquisition context. *Behaviour research and therapy*, *43*, 323-336.
- Vansteenwegen, D., Vervliet, B., Hermans, D., Beckers, T., Baeyens, F., & Eelen, P. (2006). Stronger renewal in human fear conditioning when tested with an acquisition retrieval cue than with an extinction retrieval cue. *Behaviour research and therapy*, *44*, 1717-1725.
- Vervliet, B., Vansteenwegen, D., Baeyens, F., Hermans, D., & Eelen, P. (2005). Return of fear in a human differential conditioning paradigm caused by a stimulus change after extinction. *Behaviour research and therapy*, *43*, 357-371.
- White, K., & Davey, G. C. (1989). Sensory preconditioning and UCS inflation in human 'fear' conditioning. *Behaviour research and therapy*, *27*, 161-166.

Footnotes

¹ The experiment also included a reinstatement trial. After the test phase, a non-paired US was presented. Subsequently, all participants were presented with a CS- followed by a non-reinforced CS+ trial (fixed order). The results of the reinstatement phase did not differ from the results of the test phase. It is difficult to judge whether this was due to a reinstatement effect or to the influence of the information that was given to participants before the test phase. For reasons of clarity, the reinstatement phase is not discussed any further here.